analytical chemistry

Letter

pubs.acs.org/ac

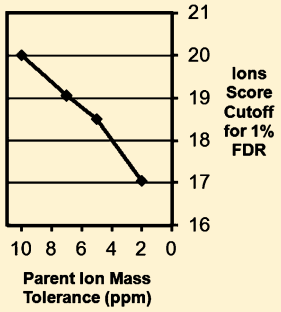
The Problem with Peptide Presumption and the Downfall of Target–Decoy False Discovery Rates

Bret Cooper*

Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, Maryland 20705, United States

Supporting Information

ABSTRACT: In proteomics, peptide-tandem mass spectrum match scores and target–decoy database derived false discovery rates (FDR) are confidence indicators describing the quality of individual and sets of tandem mass spectrum matches. A user can impose a standard by prescribing a limit to these values, equivalent to drawing a line that separates better from poorer quality matches. As a result of setting narrower parent ion mass tolerances to reflect the better resolution of modern mass spectrometers, target–decoy derived FDRs can diminish. FDRs lowered this way consequently drive down the lower-limit for peptide-spectrum match score acceptance. Hence, data quality confidence appears to improve even while fragmentation evidence for some spectra remains weak. One negative outcome can be the presumed identification of peptides that do not exist. The options researchers have to improve proteomics data confidence are not panaceas, and there may be no satisfying solution as long as peptides are identified from a circumscribed list of proteins scientists wish to find.



Last decade, amidst the promise of mass spectrometry to revolutionize biological protein discovery, proteomics researchers suddenly realized there was a problem. Peptide-spectrum match algorithms were making false matches between tandem mass spectra and candidate peptide sequences at unspecified rates, hurting the credibility of an emerging field.[1,2] In short time, however, the proteomics community rallied and developed a variety of statistical methods to model and limit false peptide-spectrum matches (PSMs).[3] One of the first programs to do so, PeptideProphet, estimated a probability of certainty for each PSM made by Sequest.[4] Around the same time, Mascot was configured to calculate an Expect value based on the total number of peptide match candidates.[5,6] This allowed the user to assess the likelihood that a PSM occurred by chance. Meanwhile, X! Tandem, another match algorithm, grew in popularity because of its own PSM confidence indicator, the Expectation value (E-value), which was based on the distribution of all matches for a spectrum.[7] As more match algorithms were developed, confidence indicators that utilized statistical and probability models became commonplace, and the means by which the scores describing PSMs were calculated were ever more innovative and varied.[3] For example, MS-GFDB creatively employed a generating function derived from combinatorics to calculate the significance of a PSM.[8,9]

Although these confidence indicators were intended to limit individual false matches, not all indicators correctly predicted the statistical rate of false matching across a set of PSMs.[10,11] A better indicator for this became known as the peptide false discovery rate (FDR), and soon scientists devised numerous ways to limit FDRs. The target–decoy method for FDR estimation became the most popular because it is easy to implement.[3] A researcher needs only to search definitive false peptide sequences (decoys) alongside target, candidate sequences. Then, to calculate the FDR, the researcher needs simply to establish a ratio of known, decoy false positive (FP) matches over all positive PSMs or set a PSM score lower limit to establish an FDR. The target–decoy method reasonably approximates FDRs, but it is not without its statistical deficiencies.[12−14] Furthermore, there is no consensus about how target–decoy should be used with each match algorithm or which "scores" are most suitable. This complicates comparisons of FDR estimates across different research applications. Thus, false data continue to nag proteomics.

More than once when looking for ways to improve data confidence, I noticed that several prominent research groups had published data sets associated with reasonably low FDRs but with PSMs having abnormally low Mascot Ions scores (16-0.1).[15] I found this unusual; I would have not expected such low scores to be associated with such low FDRs. Because abnormally low-scoring spectra have poorly discernible fragmentation ion series important for peptide sequence elucidation, I speculated that their identifications were instead predicated on parent ion mass matching.[15] I then hypothesized that as the numbers of peptide candidates are reduced through narrow parent ion mass tolerance (PIMT) parameters, thereby shifting from fragment ion to parent ion mass matching conditions, FDRs estimated by target–decoy database searching would drop, driving down the lower limit for Ions score acceptability.[15] I suspected that this would inadvertently lead to the acceptance of matches between spectra with insufficient fragmentation evidence and a few restricted peptide candidates in a narrow mass range.[15]

Received: August 23, 2012
Accepted: October 25, 2012
Published: October 29, 2012

I had hoped my hypothesis would motivate other scientists to look at this problem, especially since improvements in mass spectrometer resolution and data refinement would continue to encourage researchers to lower PIMT parameters.[16] Indeed, my essay struck a nerve. Some researchers agreed with my hypothesis about FDRs while others disagreed.[5,17] Now to affirm it, I provide data proving that target−decoy FDRs can decrease as PIMTs or numbers of peptide candidates shrink. This trend holds true for three match algorithms that employ different scoring systems. I also expose a shortcoming associated with calculating FDRs with Mascot Expect values under narrow PIMT parameters. Comfort in PSM data quality confidence remains questionable as a result.

To test my notion of decreasing target−decoy FDRs as PIMTs shrink, I used 716 088 tandem mass spectra from an *opgGH* null mutant of *Salmonella enterica* grown under osmotic stress.[18] The mutant cannot produce OpgG or OpgH, proteins required for osmotic tolerance. I will return to these deleted proteins, but let me remark that I also chose to examine *S. enterica* because of its small genome. Its derived protein sequence data file (4914 records) used for peptide-spectrum matching is more likely to lead to many of the negative effects of constrained PIMT search parameters, such as an insufficient number of candidates for the statistical modeling of false matches.[15] I then used Mascot to search the spectra under different PIMT parameters against the *S. enterica* protein sequence data file appended with the same number and lengths of statistically randomized decoy sequences (Table 1).[12] Decreasing PIMT parameters reduces the number of candidate peptide sequences considered for spectrum matching.[15] I then estimated peptide FDRs from the top-ranking matches at different Ions score limits for each altered PIMT using the eq 2Decoys/Positive_matches. Decoys is doubled to be a proxy for all FPs (all FPs = bona fide FPs + Decoys ≈ 2Decoys),[19] and Positive_matches is the sum of matches to Targets and Decoys.

An Ions score is a positive indicator of the quality of a match between a tandem mass spectrum and a candidate peptide sequence.[6] I chose it as the "score" for establishing FDRs because Mascot Ions scores were reported in all of the papers I previously examined.[15] Another Mascot score, the Expect value, did not appear to be a factor in those papers, at least not for FDR calculations. Thus, my original argument was based on the use of Ions scores for calculating FDRs, and my first intention was to test that. As presented in Table 1 for a 10 ppm PIMT Mascot search, the FDR progressively grew as the Ions score limit was reduced from 25 to 10 (horizontal 10 ppm data). This is expected because, as Ions scores decrease, the match quality gets worse and false matching increases. The same trend occurred at lower PIMTs, although lowering the PIMT reduced the number of peptide candidates from the next highest PIMT parameter and reduced the total number positive matches (vertical positive matches data in Table 1). This is likely due to some peptide candidates falling out of consideration. Again, this is expected since the PIMT parameter is directly related to the number of peptide sequence candidates.[15]

The disturbing observation from this data, however, is that, at any given Ions score cutoff, the FDR decreased as PIMT dropped (vertical FDR data in Table 1). This means that, as PIMT decreased, decoy matches fell out of consideration faster than all positive matches (or all positive matches did not disappear as fast as decoys). This trend for decreasing FDRs with each PIMT drop held true even when the decoy database

Table 1. Numbers of Positive Matches, Decoy Matches, and Estimated FDRs per Mascot Ions Score Lower Limit for Tandem Mass Spectra from *S. enterica* under the Given Parent Ion Mass Tolerance Parameters[a]

parent ion mass tolerance, ppm	type of match	matches with Mascot Ions score ≥25	FDR (Mascot Ions score ≥25), %	matches with Mascot Ions score ≥20	FDR (Mascot Ions score ≥20), %	matches with Mascot Ions score ≥15	FDR (Mascot Ions score ≥15), %	matches with Mascot Ions score ≥13	FDR (Mascot Ions score ≥13), %	matches with Mascot Ions score ≥10	FDR (Mascot Ions score ≥10), %
10	decoys	159		469		1748		3170		7900	
	positive matches	83 076	0.38	94 889	0.99	109 022	3.2	116 430	5.4	132 539	11.9
7	decoys	139		372		1361		2436		5987	
	positive matches	81 786	0.34	93 127	0.88	106 317	2.6	112 874	4.3	126 367	9.5
5	decoys	131		325		1100		1909		4624	
	positive matches	80 354	0.32	91 310	0.71	103 702	2.1	109 606	3.5	121 250	7.6
2	decoys	100		214		601		975		2259	
	positive matches	66 715	0.30	75 314	0.57	84 582	1.4	88 664	2.2	96 082	4.7

[a]Other Mascot version 2.3.02 search parameters were ±0.8 Da fragment ion mass tolerance, tryptic digest, static carboxyamidomethylation for C, monoisotopic mass values, 1 missed cleavage, and C#13=1. Ions scores, Expect values, and qmatch values were obtained from the Mascot results file. The median parent ion mass deviation for this data set is less than 2 ppm.[18]

dx.doi.org/10.1021/ac303051s | *Anal. Chem.* 2012, 84, 9663−9667

Table 2. Estimated FDRs per Mascot Expect Upper Limit for Tandem Mass Spectra from Table 1 under the Same Search Conditions[a]

parent ion mass tolerance, ppm	FDR (Expect ≤0.01), %[b]	matches with Identity scores of 13 (Expect ≤0.01), %	FDR (Expect ≤0.05), %[c]	matches with Identity scores of 13 (Expect ≤0.05), %	FDR (Expect ≤0.075), %	FDR (Expect ≤0.1), %	FDR (Expect ≤0.5), %
10	0.52	**12**	1.8	**12**	2.8	4.2	23
7	0.53	**22**	2.0	**22**	3.1	4.4	25
5	0.57	**37**	2.1	**35**	3.4	4.7	25
2	0.54	**97**	2.0	**97**	3.1	4.3	22

[a]FDR stays consistent as parent ion mass tolerance drops at any given Expect value limit, but the percentage of matches with Identity scores of 13 rises (in bold). [b]Smallest Ions score is 19.79. [c]Smallest Ions score is 12.97.

Table 3. Top-Ranking Peptide Sequences, Ions, Identity and Expect Scores, and Numbers of Candidates Considered (qmatch) for Orbitrap-LTQ Tandem Mass Spectra from an *S. enterica opgGH* Mutant Grown in Low Osmotic Medium[a]

protein	peptide match	Mascot Ions score (10 ppm)	Mascot Identity score (10 ppm)	Mascot Expect (10 ppm)	Mascot qmatch (10 ppm)	Mascot Ions score (2 ppm)	Mascot Expect (2 ppm)	Mascot Identity score (2 ppm)	Mascot qmatch (2 ppm)
OpgG									
	GWRLMLR + Ox(M)	12.59	16.81	0.13	48	12.59	0.055[b]	7.78[c]	6
	DLGFAGFKVLYPINSK	10.67	13.80	0.1	24	10.67	0.086[b]	4.77[c]	3
	MRWLGAAIMLTLYASSSWAFSIDDVAK + Ox(M)	10.24	14.15	0.12	26	no match	no match	8.45[c]	7

[a]The mutant does not produce OpgG, so the peptide matches are false. [b]Thresholding implemented. [c]Nonthreshold calculation.

was created by reversing sequences between R/K characters (Supplementary Table 1, Supporting Information). The trend for decreasing FDRs with each PIMT drop was also true when X! Tandem and MS-GFDB were used instead of Mascot, and the respective PSM scores, E-values and p-values, were used to estimate the FDRs at each limit (Supplementary Table 2, Supporting Information).

Thus, these data support my original hypothesis that target−decoy FDRs decrease as the number of peptide candidates grows smaller. Nevertheless, Cottrell and Creasy previously disputed my concept, at least with respect to Mascot results.[5] Their reasoning was that if the quality of matching becomes worse as scoring decreases (as the number of candidates gets smaller), FDRs must get higher, not lower. Our views appear contradictory, so I estimated FDRs using their recommended score, the Mascot Expect, instead of the Ions score; but first, a little background: Default Expect is pseudocoded as $0.05 \times \text{qmatch} \times 10^{(-\text{Ions score}/10)}$ where qmatch is the number of peptide candidates within a prescribed PIMT. Hence, the equation relates fragment ion matching quality to the number of peptide candidates. Substituting qmatch with its equivalent, $10^{(\text{Identity score}/10)}$, allows direct comparison between the Identity and Ions scores: When they are equal, Expect is 0.05. According to the Mascot model, as Expect diminishes, there is improved confidence that a match to any particular peptide is not random. Note, however, that the model suffers when the Identity score is less than 13 since it would no longer be theoretically possible to obtain at least 1-out-of-20 peptide match confidence (i.e., Expect = 0.05). Therefore, Mascot thresholds the Identity score to 13 when this condition is met.[5] I will show the ramifications of thresholding in the next paragraphs. In the meantime, note it is common practice to accept a peptide with an Expect less than or equal to 0.05, although a researcher could prescribe an alternative cutoff. Expect is calculated per match; no single Expect should extrapolate to a set of different peptide matches.

The data in Table 2 show at the tested Expect cut-offs that there was no appreciable increase in FDRs at decreasing PIMTs. In fact, the FDRs remained consistent. Even for separate target and decoy database searches and the appropriate FDR calculation correction,[19] this trend held true (not shown). Therefore, Cottrell's and Creasy's prediction of *rising* FDRs is not evident, even when using Expect values. Notwithstanding, they could argue that it is better to use the Expect to estimate FDRs because of the consistency of the FDRs across a series of PIMTs. That thought, however, would be premature. At an Expect cutoff of 0.01 for a 10 ppm PIMT, which gave an FDR of 0.52%, the lowest Ions score for any spectrum was 19.79 (Table 2). Meanwhile, with the Ions score method to achieve a similarly low FDR of 0.38% at the same PIMT, the lowest Ions score for any spectrum was 25 (Table 1). This means that more poorer-scoring spectra, in terms of Ions score magnitude, were among the true positives when Expect values were used for calculating FDRs than when Ions scores were used. It is difficult to build an argument that justifies the use of Expect values for FDR calculations when it results in the inclusion of more poorer-scoring spectra. Worse still, as PIMTs dropped from 10 to 2 ppm for a constant (and standard) 0.05 Expect value cutoff, the percentage of all positive matches with Identity scores of 13 rose from 12% to 97% (Table 2). This means that a narrow 2 ppm search reduced the numbers of candidates to the point that it was no longer possible to ascertain 1-out-of-20 match confidence for 97% of all positive matching spectra. Consequently, the Expect values were merely the Ions scores in nearly all cases, and therefore, no longer reflected the true number of trials. Hence, it is not fair to view FDRs between the 10 ppm PIMT and the 2 ppm PIMT searches in Table 2 as being consistent because the Expect values have inconsistent mathematical meanings. This makes it difficult to assess whether FDRs are static or change for better or worse when using Expect values and narrow PIMTs. More than likely though, the artificiality of the Mascot thresholding masks the true FDR drop, the type seen in Table 1 and Supplementary Tables 1 and 2, Supporting Information.

Despite the Expect value breakdown, it might not be apparent to a casual reader what is wrong with lowering the

dx.doi.org/10.1021/ac303051s | *Anal. Chem.* 2012, 84, 9663−9667

PIMT to get a lower FDR, especially since a low FDR is the modern gold standard in proteomics. To understand, let us track some specific spectra falsely matched to peptides such as the false matches to the OpgG protein since the *S. enterica* mutant does not produce it (Table 3).[18] At a 10 ppm PIMT and a 5% FDR, these bona fide falsely matched peptides were not included among the positively matched peptides in Table 1. However, after shifting the PIMT down to 2 ppm, but maintaining the same 5% FDR as before, the Ions score cutoff dropped below 10, which means that 2 of the 3 falsely matched OpgG peptides moved from the prior false positive data set to the positive match data set (compare Tables 1 and 3). Therefore, in this case, parent ion mass selectivity and target−decoy FDRs worked against each other to engender these low-scoring FP peptide matches, hence one problem with lowering the FDR by constraining PIMT. Yes, it is possible to limit the FDR to 1% to prevent the movement of the FP OpgG peptides to the true positive data set, but this is knowledge after the fact. A user should beware that changing the FDR bar is only an apparent fix because similar circumstances will arise for other peptides wherever an arbitrary line is drawn.

It is possible to counter the previous example and contend that it is meaningless to consider these matches in the first place because they do not pass the standard Expect cutoff of 0.05, but this plea ignores some simple facts and overlooks the point the example makes. First, the Expect value is subject to whimsical manipulation of search parameters such as PIMT, the numbers of tryptic ends, missed cleavages, and variable mass modifications. Because these parameters affect the number of peptide sequence candidates, it is all too easy to find a favorable setting to achieve desirable match confidence. Second, a scientist accepting such low-scoring peptides has probably done just that: ignored the Expect value and only considered the Ions score. Therefore, under the extreme conditions provided by the example, it is entirely possible to arrive at a situation where the false matches to the OpgG peptides are considered real.

So what are the options? For Mascot users, there are basic guidelines to follow: If the Mascot Ions and Identity score relationship is no longer meaningful as these scores approach zero, as I proved before,[15] and if the Expect value scoring system mathematically deteriorates near 13, then peptide matches with Ions scores below 13 are sufficiently doubtful. There may be no good reason to consider any lower Ions scores.

As for target−decoy, a common way to reduce the number of false positives is to lower the FDR limit. Unfortunately, this may lead to unnecessary data waste. Another possibility is to search more peptide candidates. For Mascot, this would prevent the Expect value accuracy from nose-diving. Ironically, widening the PIMT to accomplish this belies the benefit of high-resolution, high-mass accuracy mass spectrometry. Widening the search space by adding more decoys to the protein sequence data file also betrays the simple principle of estimating FDRs using the standard target−decoy approach.

Of course, if the identification of proteins were the overall goal, a protein level criterion may help. For example, Protein Prophet, PANORAMICS, and other programs calculate a protein identification probability from PSM probabilities.[20−24] On the basis of the OpgG peptides in Table 3, the PANORAMICS[2] probability for the OpgG protein being found was <10% for the 10 ppm search and 50.4% for the 2 ppm search.[18] Thus, while some of the PSM data in Table 3 may be construed to imply the OpgG protein was identified, the additional level of discrimination provided by the protein probability model would imply that the protein was not reliably found. Hence, the OpgG peptides would be disregarded accordingly. One advantage of protein-probability models is that they can be independent of decoy database searching,[20−24] but the user should beware of models that rely on peptide target−decoy FDRs to estimate protein probabilities for the reasons presented here.[25−27] Note, narrowing the PIMT may also manipulate protein probabilities, but this additional level of discrimination may serve as a buffer against the peptide-level effects shown here.

In the end, there may be no remedy. This is partly because peptide identification by tandem mass spectrometry becomes inherently biased as soon as we set out to find the proteins already on our list. Any affirming PSM score to a protein on that list can give some reason to believe that it exists, even if it is experimentally precluded. This brings to mind Pascal's Wager for God, where probability dictates that it is better to wager for existence than not: that is, until one accepts that no probability of existence is a logical option.[28] The predetermined expectation of protein discovery is the foundation for peptide presumption and is the root of our chase to eliminate FPs. We must deal with this; otherwise, poor-scoring PSMs could have a stronger likelihood of being improbable than what the prescribed confidence measures should indicate or allow.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Address: 10300 Baltimore Ave., Bldg. 006, Rm. 213, Beltsville, MD 20705. Phone: 301-504-9892. E-mail: bret.cooper@ars.usda.gov.

Notes

The author declares no competing financial interest.

■ REFERENCES

(1) Patterson, S. D. *Nat. Biotechnol.* **2003**, *21*, 221−222.
(2) White, F. M. *Sci. Signal* **2011**, *4*, 8.
(3) Nesvizhskii, A. I. *J. Proteomics* **2010**, *73*, 2092−2123.
(4) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383−5392.
(5) Cottrell, J. S.; Creasy, D. M. *J. Proteome Res.* **2011**, *10*, 5272−5273.
(6) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(7) Fenyo, D.; Beavis, R. C. *Anal. Chem.* **2003**, *75*, 768−774.
(8) Kim, S.; Gupta, N.; Pevzner, P. A. *J. Proteome Res.* **2008**, *7*, 3354−3363.
(9) Kim, S.; Mischerikow, N.; Bandeira, N.; Navarro, J. D.; Wich, L.; Mohammed, S.; Heck, A. J.; Pevzner, P. A. *Mol. Cell. Proteomics* **2010**, *9*, 2840−2852.
(10) Kall, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. *J. Proteome Res.* **2008**, *7*, 40−44.
(11) Storey, J. D.; Tibshirani, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440−9445.
(12) Feng, J.; Naiman, D. Q.; Cooper, B. *Bioinformatics* **2007**, *23*, 2210−2217.
(13) Gupta, N.; Bandeira, N.; Keich, U.; Pevzner, P. A. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1111−1120.

(14) Barboza, R.; Cociorva, D.; Xu, T.; Barbosa, V. C.; Perales, J.; Valente, R. H.; Franca, F. M.; Yates, J. R., 3rd; Carvalho, P. C. *Proteomics* **2011**, *11*, 4105−4108.
(15) Cooper, B. *J. Proteome Res.* **2011**, *10*, 1432−1435.
(16) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18132−18138.
(17) Yaffe, M. B. *F1000 Prime*, 2011. Website: http://f1000.com/prime/9305956#eval9918054. Accessed 11/2012.
(18) Cooper, B.; Chen, R.; Garrett, W. M.; Murphy, C.; Chang, C.; Tucker, M. L.; Bhagwat, A. A. *J. Proteome Res.* **2012**, *11*, 1720−1727.
(19) Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4*, 207−214.
(20) Feng, J.; Garrett, W. M.; Naiman, D. Q.; Cooper, B. *J. Proteome Res.* **2009**, *8*, 5396−5405.
(21) Feng, J.; Naiman, D. Q.; Cooper, B. *Anal. Chem.* **2007**, *79*, 3901−3911.
(22) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 4646−4658.
(23) Gerster, S.; Qeli, E.; Ahrens, C. H.; Buhlmann, P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12101−12106.
(24) Gupta, N.; Pevzner, P. A. *J. Proteome Res.* **2009**, *8*, 4173−4181.
(25) Spirin, V.; Shpunt, A.; Seebacher, J.; Gentzel, M.; Shevchenko, A.; Gygi, S.; Sunyaev, S. *Bioinformatics* **2011**, *27*, 1128−1134.
(26) Choi, H.; Ghosh, D.; Nesvizhskii, A. I. *J. Proteome Res.* **2008**, *7*, 286−292.
(27) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(28) Seife, C. *Zero: The Biography of a Dangerous Idea*; Penguin Books: New York, NY, 2000; p 248.